

With respect to prior managerial experience, Dr. Ward analyzed Dr. Lundquist's application coding data to determine the relationship between prior management experience and sales productivity. He concluded that actual first year sales production is significantly increased with more management experience.¹⁶¹

Dr. Ward also faults Drs. Lanier and Lundquist for failing to consider evidence of prior sales volume in their analyses. Analyzing the WRG data, Dr. Ward concluded that the most important factor in predicting subsequent sales productivity at Sterling is the volume of sales achieved by applicants before hire.¹⁶² With respect to Dr. Lundquist's observation that sales volume was not in fact "proven," Dr. Ward "found ample evidence that sales volumes are documented."¹⁶³

Dr. Ward also criticizes Dr. Lundquist's conclusions regarding coder reliability and consistency, arguing she should have examined separately for each type of experience, consistency should have been evaluated on a question-by-question basis, and concluding that calculating "match rates" on this basis reveals significant disagreement among coders.¹⁶⁴

With respect to Dr. Lanier's statistical analyses, Dr. Ward challenges Dr. Lanier's conclusion that the WRG is "a poor predictor of sales productivity." Dr. Ward found that "as the WRG wage goes up by one dollar, average annual sales increased by approximately \$32,207."¹⁶⁵

Finally, Dr. Ward contends that Dr. Lanier committed a number of methodological and data errors that call his conclusions into question.¹⁶⁶

¹⁶¹ *Id.* at 2. Dr. Ward also questions Dr. Lundquist's opinions regarding the five-year cap on experience and the failure to include Direct Customer Service experience. Analyzing Dr. Lundquist's data, Dr. Ward found that experience over five years had no significant impact on first year sales production and that Direct Customer Service experience is not associated with higher sales. *Id.*

¹⁶² *Id.* at 3, 7, 10-11.

¹⁶³ *Id.* at 3.

¹⁶⁴ *Id.* at 3-4, 18-20.

¹⁶⁵ *Id.* at 4, 6-7 and n.8.

Dr. Lundquist's report with respect to compensation is also addressed by Sterling's experts Drs. Dunleavy and Sady, each of whom has a Ph.D. in I/O Psychology. Drs. Dunleavy and Sady contend that Dr. Lundquist's use and analysis of application data and conclusions regarding job-relatedness are unscientific, unreliable and flawed.

First, Drs. Dunleavy and Sady contend that Dr. Lundquist used unreliable and unscientific methods in coding applications and made no scientific attempt to measure the accuracy of coding, resulting in significant coding errors, and that the sub-study designed to verify that data was consistently coded was flawed.¹⁶⁷

Second, Drs. Dunleavy and Sady criticize Dr. Lundquist's opinions regarding job-relatedness, contending that the research literature supports a link between managerial experience and sales productivity, and that Dr. Lundquist failed to evaluate all of the available and relevant information in O*NET.¹⁶⁸ Using an expanded universe of relevant worker characteristics, Sterling's experts conducted an analysis of worker characteristics necessary for two O*NET positions--Retail Salespersons (considered by Dr. Lundquist), as well as First-Line Supervisors of Retail Sales Workers, concluding that managerial experience is job-related for retail sales jobs.¹⁶⁹

Drs. Dunleavy and Sady also criticize Dr. Lundquist's conclusion that the WRG was inconsistently applied and therefore unreliable.¹⁷⁰ They state that the research literature on HR process structure "suggests that semi-structured tools with guardrails on content and process

¹⁶⁶ *Id.* at 5, Appendix 2.

¹⁶⁷ Report of Sterling experts Drs. Eric M. Dunleavy and Kayo Sady (Dunleavy/Sady Report) at 25-27.

¹⁶⁸ *Id.* at 12-18.

¹⁶⁹ *Id.* at 19. Drs. Dunleavy and Sady also point out that Dr. Lundquist's opinions regarding job-relatedness are contradicted by the statistical analysis of WRG data performed by Sterling's expert Dr. Ward. *Id.* at 18-19.

¹⁷⁰ Dunleavy/Sady Report at 19-24.

reduce measurement error and vulnerability to EEO bias.”¹⁷¹ They assert that the WRG process imposes structure on content and process features that reduces the likelihood of both measurement error and vulnerability to EEO bias, and observe that although Dr. Lundquist asserts that the WRG may be “vulnerable” to rating biases, there is no evidence in her report related to the extent to which bias actually occurred.¹⁷² With respect to job-relatedness, Drs. Dunleavy and Sady note that although the WRG was not based on a formal job analysis, “it was developed and refined by subject matter experts over an extended period of time involving multiple development stages and reviews,” using “internal job performance metrics to determine prior experience factors associated with sales performance.”¹⁷³

Finally, with respect to starting pay, Sterling offers the testimony of Dr. Stockdale. Dr. Stockdale holds a Ph.D. in I/O Psychology, is Professor and Chair of Psychology at Indiana University-Purdue University, has for over 25 years conducted both field and laboratory research on gender discrimination and sexual harassment, has authored numerous publications, and has previously testified as an expert witness in several cases involving allegations of sex discrimination.¹⁷⁴

¹⁷¹ *Id.* at 22.

¹⁷² *Id.* at 21-22.

¹⁷³ *Id.* at 24. Drs. Dunleavy and Sady also challenge Dr. Lundquist’s assertion that the WRG cap of five years of experience is problematic because it reduces job-relatedness and may increase adverse impact against women. They argue that use of an experience cap is not unreasonable because (1) applicant self-report data from applications may be less accurate for older experience; and (2) the strength of the relationship between years of experience and job performance may diminish over time, such that years of experience no longer correlates with productivity past a certain number of years. *Id.* at 25. They further note that Dr. Ward’s statistical analysis showed that the five-year experience cap impacts men and women about the same, and that experience beyond five years has no statistically significant impact on productivity. *Id.*

¹⁷⁴ Report of Sterling’s expert Dr. Margaret S. Stockdale (Stockdale Report) at 5-6 and Exhibit A (Curriculum Vitae).

Dr. Stockdale opines that that Sterling's corporate practices, including the WRG, "provide minimal opportunity for managerial discretion in determining starting pay."¹⁷⁵ Dr. Stockdale also discusses the literature on "gender role socialization," as well as national occupational statistics, from which she concludes that men are likely to have more managerial experience than women, which provides an explanation for gender differences in starting pay.¹⁷⁶

Merit Pay

Regarding merit pay, Dr. Ward opines that "women receive higher unadjusted pay raises than men."¹⁷⁷ He does not, however, directly address the conclusion of Claimants' experts that the merit pay system perpetuates initial pay disparities that are based on the use of prior experience factors that Claimants contend are not job-related, and that the merit pay system fails to remedy compensation gaps between male and female employees with equal performance.

Promotion

Dr. Ward's report contains an extensive critique of Dr. Lanier's analysis of promotions, finding numerous data errors and anomalies.¹⁷⁸ Dr. Ward's most significant objection to Dr. Lanier's analysis is that Dr. Lanier largely ignored information contained in the CAR system with respect to interest in being promoted and willingness to relocate to receive a promotion. Dr. Ward concludes that if the pool of eligible candidates is limited to those who registered their interest and availability in CAR, the promotion rate for women equals or exceeds the promotion

¹⁷⁵ Stockdale Report at 7.

¹⁷⁶ *Id.*

¹⁷⁷ Ward Report at 5, 7.

¹⁷⁸ Dr. Ward finds that "Dr. Lanier's data errors cause him to find promotional differences where there are none. Approximately 73% of sales employees are women and approximately 72% of Assistant Managers are women. There is no gender difference in promotion to this first level of management." Ward Report at 7-8.

rate of men.¹⁷⁹ Accordingly, the critical difference between Dr. Lanier and Dr. Ward with respect to disparity in promotion rate is whether the pool of available candidates should be defined by those who registered their interest in CAR. Dr. Ward performed a number of analyses to determine whether the alleged improprieties in the administration of CAR disadvantaged women. Specifically, Dr. Ward addressed Dr. Lanier's concern that because there are many promotions from the CAR register that occur soon after, or even concurrently with registration, a promotional "tap on the shoulder" may have resulted in CAR registration. Dr. Ward concluded that if there are "taps," they are more often on women.¹⁸⁰

Dr. Ward also faults Dr. Lanier's analysis of disparities in the time it took for male and female employees to get promoted at Sterling. Specifically, Dr. Ward found that after accounting for the timing of expression of interest (as determined from the CAR data), "women are promoted slightly faster than men" (not statistically significant).¹⁸¹

Finally, with respect to Sterling's promotion practices, Sterling offers the expert opinion of Dr. Stockdale, who concludes that based upon "gender role socialization" and national occupational statistics, men are more interested in promotion than women, which is consistent with their proportionally greater representation in CAR.¹⁸² Dr. Stockdale also endorses Sterling's promotion practices and performance appraisal process, concluding that Sterling "relies heavily on objective, results-oriented criteria" that are "not vague and leave little room for interpretation."¹⁸³ Dr. Stockdale faults Dr. Outtz for "cherry-picking" Sterling's promotion

¹⁷⁹ CAR was introduced in April 2007. Dr. Ward opines that the conclusions reached from an analysis of the CAR period can be extrapolated to the full 2003-2012 period, based upon an analysis of all promotions in that period, which showed that the percent of promotions that go to women in the full period is very similar to the percentages in the CAR period. Ward Report at 36.

¹⁸⁰ *Id.* at 21-22.

¹⁸¹ *Id.* at 5, 21-22 and Table 5.

¹⁸² Stockdale Report at 7-8.

¹⁸³ *Id.* at 8.

criteria, and asserts that his analysis of Sterling's promotion practices is "not grounded in scientific methods nor in the established practices of I/O psychologists."¹⁸⁴

Extent of Disparities

With respect to Dr. Lanier's conclusion that gender differences in pay and promotion are widespread across districts (pay) or regions (promotion), Dr. Ward points out that Dr. Lanier offered no analysis to demonstrate whether or where the differences are statistically significant. According to Dr. Ward's analysis, the pay differentials identified by Dr. Lanier vary by district, and promotion shortfalls vary by region, sometimes favoring women and sometimes favoring men.¹⁸⁵

Compensation and Promotion: Claimants' Rebuttal

The critique by Sterling's experts regarding Claimants' analysis of compensation and promotion is addressed in the rebuttal reports of Drs. Lanier and Lundquist.

Dr. Lanier's rebuttal report begins by pointing out that he and Dr. Ward both observe the same gender-related pay gaps, and "agree that the pay difference between male and female employees begins with the assignment of initial pay on hire; that Sterling's system of merit pay raises perpetuates, rather than eliminate, the gender pay gaps that were created at hire; that pay changes associated with promotions to management also perpetuate the pay differences; and that females out performed males, on average, with respect to performance evaluation scores."¹⁸⁶ Their analyses differ based upon whether these gaps are justified by consideration of prior experience factors that Claimants contend are not job-related. As Dr. Ward testified, "[A]ll of

¹⁸⁴ *Id.*

¹⁸⁵ Ward Report at A2-19; Tables 20 and 21.

¹⁸⁶ Rebuttal Expert Report of Dr. Louis R. Lanier (Lanier Rebuttal Report) at 1.

the pay differences up the line that you see, all those gender pay differences, are due to starting pay.”¹⁸⁷

Dr. Lanier notes that Dr. Lundquist concludes in her rebuttal analyses that the only valid predictor of first-year sales is prior personal jewelry sales volume.¹⁸⁸ Dr. Lanier observes that the use of volume data in setting initial pay is optional for managers, even under the WRG, and that among applicants with prior sales experience, men are more likely to have recorded sales volume data than are women.¹⁸⁹ Dr. Lanier provides an analysis of the WRG timeframe showing that if only prior personal jewelry sales volumes are included, there are significant differences in starting pay, adverse to women for both part-time and full-time employees.¹⁹⁰

Dr. Lanier also faults Dr. Ward for relying on measures of employee sales performance without controlling for the store in which they work, despite Dr. Ward’s acknowledgement of the importance of stores at which employees work in affecting the volume of sales they are able to make.¹⁹¹

With respect to Dr. Ward’s criticism of his data errors, Dr. Lanier includes rebuttal analyses that use Dr. Ward’s data, and which for the most part do not materially change Dr. Lanier’s earlier findings and conclusions.¹⁹²

¹⁸⁷ *Id.* ¶5 (citing Ward Dep. at 244:18-21).

¹⁸⁸ *Id.* ¶10.

¹⁸⁹ Lanier Rebuttal Report at 8, n. 3.

¹⁹⁰ *Id.* ¶9.

¹⁹¹ *Id.* ¶¶2(d); 12.

¹⁹² *Id.* ¶2(f) and Tables B3 to B15.

Regarding promotions, Dr. Lanier criticizes Dr. Ward for overestimating the impact of an employees' expression of willingness to relocate in CAR, and generally reiterates concerns regarding the validity of CAR as a measure of female interest in promotion, noting that Dr. Ward provides no empirical analysis of promotions prior to CAR.¹⁹³

Dr. Lundquist's rebuttal report provides an analysis of Dr. Ward's statistical findings with respect to compensation, noting at the outset that none of Sterling's experts offers an opinion regarding gender differences in pay prior to the WRG period. Most importantly, Dr. Lundquist points out that Dr. Ward failed to control for store sales volume, *i.e.*, opportunity to sell, when analyzing the relationship between experience and sales productivity.¹⁹⁴ Dr. Lundquist reanalyzed Dr. Ward's data using a hierarchical regression analysis.¹⁹⁵ She concluded that, controlling for store volume, "the previous experience variable in the WRG with the most predictive power was jewelry Personal Sales volume, uniquely accounting for approximately 3% of the variance in annualized sales (squared semi-partial correlation of 0.028) for full-time employees and 1% for part-time employees."¹⁹⁶ Two other statistically significant WRG variables (non-jewelry personal sales volume for full-time employees and other management-jewelry volume for part-time employees) "were much less useful, each uniquely accounting for 0.1% of the variance in predicted sales performance."¹⁹⁷ She found that "Store Management experience and volumes do not predict sales performance and Other Management volume only predicts a very small percent of the variance in sales performance and only for part-time

¹⁹³ *Id.* ¶¶26-31,

¹⁹⁴ Lundquist Rebuttal Report at 15-16; Dr. Lundquist also faults Dr. Ward for failing to separately analyze full-time and part-time employees. *Id.* at 17-18.

¹⁹⁵ *Id.* at 17-20.

¹⁹⁶ *Id.* at 20 and Table 2.

¹⁹⁷ *Id.*

employees.”¹⁹⁸ She concluded that management experience has “minimal” value in predicting success as a sales associate, and should not have been credited in setting starting pay rates.¹⁹⁹

Dr. Lundquist also examined the pattern of gender differences for all of the experience and volume variables in the WRG. She found statistically significant gender differences in variables that are not predictive of job performance, *e.g.*, store management experience, which would disproportionately credit men for experience that is not predictive of success on the job.²⁰⁰

Dr. Lundquist also analyzed the reliability of the WRG process for setting starting salaries. She concluded that DMs failed to follow company guidelines for almost 24% of hires in the WRG timeframe.²⁰¹ Dr. Lundquist also examined the extent to which the DMs’ entry of experience data into the WRG was consistent with her coders’ entry of experience data, finding statistically significant differences in the coding of sales experience.²⁰² Dr. Lundquist also found inconsistencies in the verification of sales volume.²⁰³ Dr. Lundquist concluded that DMs “may have been manipulating their WRG entries to obtain the desired wage rate.”²⁰⁴ Dr. Lundquist further noted that men are more likely to have two or more WRG records than women, which she views as evidence that DMs are “shopping” for wages more frequently for men than for women.²⁰⁵ Dr. Lundquist concludes that the exceptions to guidelines and inconsistencies in the implementation of Sterling’s WRG process “undoubtedly compromised the accuracy and job-relatedness of the company’s starting pay decisions.”²⁰⁶

¹⁹⁸ *Id.*

¹⁹⁹ *Id.* at 19.

²⁰⁰ *Id.* at 21-22.

²⁰¹ *Id.* at 24-25.

²⁰² *Id.* at 26-27.

²⁰³ *Id.* at 27-28.

²⁰⁴ *Id.* at 27.

²⁰⁵ *Id.* at 25-26.

²⁰⁶ *Id.* at 28. Dr. Lundquist does not provide an analysis of whether the coding and verification inconsistencies adversely affected women.

Dr. Lundquist notes that Sterling's experts portray Sterling's processes as "formulaic" and "objective" without any investigation of their actual use, and that none of Sterling's experts can testify as to the consistency and reliability of the information entered into the WRG.²⁰⁷ Dr. Lundquist points out that Dr. Dunleavy's statement that the WRG process is "unequivocally" valid ignores the professional requirements for establishing validity, and that Dr. Dunleavy admitted that he did not attempt to review the job descriptions for the key positions in this case.²⁰⁸

Overall, Dr. Lundquist concludes that Sterling's experts have not presented evidence supporting the job relatedness of the WRG.

With respect to Sterling's critique of her applicant coding study, Dr. Lundquist contends that "the methods and principles used to code applications in this matter are regularly used by Industrial/Organizational Psychologists to evaluate applicants' training and experience in both research and practice."²⁰⁹ She notes that the coding procedure used in her study was similar to the Minimum Qualifications and Training & Evaluation assessments used routinely by employers to screen and evaluate applicants, and that the methods and principles used in her study are also consistent with the research methodology known as content analysis.²¹⁰ With respect to Dr. Ward's argument that Dr. Lundquist should have computed intraclass correlations

²⁰⁷ *Id.* at 3, 6, 7, 29-30 (Dunleavy testified that he did not examine and has no opinion about whether the actual compensation practices, *i.e.*, the information gathered and entered into the WRG, were implemented in a reliable or consistent manner, and that contrary to the statement in his report that the WRG was subject to internal audit, he was unable to identify any evidence that Sterling actually conducted internal audits of the WRG process; Dr. Stockdale testified that assessing the reliability of the information input to the WRG was outside the scope of her assignment; Dr. Ward testified that he did not study how managers code the inputs to the WRG).

²⁰⁸ *Id.* at 9-14 (*citing* SIOP Principles, and noting that Dr. Dunleavy relies on the correlations of Dr. Ward for evidence of validity "without meeting the basic requirements for research and documentation of validity evidence").

²⁰⁹ *Id.* at 31.

²¹⁰ *Id.* at 31-32 (*citing* authorities).

(ICCs) by experience category, rather than across all categories of coded experience, Dr. Lundquist explains that Dr. Ward's approach is contrary to her objective to represent the level of consistency in coders' overall judgments regarding the entirety of the applicants' experience as reflected on the application form, since it is Sterling's judgment across the various experience items that leads to an overall decision about an applicant's starting pay.²¹¹ She notes that Dr. Dunleavy reviewed her single measure ICCs and conceded that they were more than sufficient to indicate the reliability of the coding study.²¹² She further notes that Dr. Ward's suggested alternative (matched inter-coder values) is unsupported in either the research literature or professional practice, and observes that Dr. Stockdale uses ICCs to establish the consistency of her coders in content analyzing comments on Sterling's employee surveys.²¹³

In response to the new analyses conducted by Dr. Lundquist, Sterling has submitted a declaration by Dr. Ward.²¹⁴ With respect to the asserted flaws in his conclusion that WRG inputs are statistically important predictors of subsequent sales, Dr. Ward contends that "each and every statistical objection made by Dr. Lundquist is invalid."²¹⁵ With respect to Dr. Ward's asserted failure to control for store volume or "opportunity to sell," Dr. Ward argues that Dr. Lundquist's use of store volume "violates fundamental assumptions used in regression analysis," and fails to answer the question of whether new employees with more of the WRG inputs sell more than new employees with fewer of those inputs within the same store.²¹⁶ Dr. Ward evaluated the impact of

²¹¹ *Id.* at 32.

²¹² *Id.* at 33.

²¹³ *Id.* at 33-34.

²¹⁴ Declaration of Michael P. Ward, dated February 14, 2014 (Ward Declaration) (Exhibit 3 to Sterling's Reply in Support of its Motion to Strike the Report and Testimony of Dr. Lundquist.)

²¹⁵ Ward Declaration ¶3.

²¹⁶ *Id.* ¶¶ 9-10.

the WRG inputs within stores, concluding that the WRG inputs, including manager experience, remain statistically important predictors of subsequent sales.²¹⁷

Moreover, Dr. Ward contends that Dr. Lundquist's own regression shows that "prior store manager experience" and "other manager prior experience" are statistically significant predictors of subsequent sales. Specifically, Dr. Ward points out that Dr. Lundquist's claim that management experience factors are *not* statistically significant in explaining sales of full-time Sales Associates is based upon Dr. Lundquist's adoption of a higher standard for statistical significance than she uses in her prior reports: instead of two standard deviations, or a probability of less than 5%, Dr. Lundquist requires a probability of below 0.01. Using two standard deviations, Dr. Lundquist's analysis shows that both the "other management" and "store management" experience variables are statistically significant for full-time Sales Associates (standard deviations of 2.45 and 2.19, respectively).²¹⁸ Dr. Ward also faults Dr. Lundquist's use of "hierarchical regression," because by simply reversing the order of inclusion of the variables, the conclusion Dr. Lundquist attempts to draw about their relative importance changes completely (e.g., if store manager experience is introduced first, it becomes statistically significant).²¹⁹

With respect to Dr. Lundquist's claim that Sterling's use of the WRG was inconsistent and that almost 24% of the hires were subject to "exceptions," Dr. Ward concludes that "this finding results almost entirely from Dr. Lundquist's lack of understanding of Sterling's policy and data practices," and that correcting the errors in her analysis reduces the "exception" rate to 3.5%.²²⁰ Moreover, Dr. Ward finds that these exceptions *advantaged* women by twelve cents per

²¹⁷ *Id.* ¶¶28-32.

²¹⁸ *Id.* ¶¶11-14.

²¹⁹ *Id.* ¶¶18-20.

²²⁰ *Id.* ¶4.

hour relative to men (statistically insignificant).²²¹ With respect to Dr. Lundquist's finding that men are more likely to have two or more WRG records than women (which she interprets as evidence that DMs are "shopping" for wages more frequently for men than for women), Dr. Ward found that this was explained entirely by the fact that men are more likely to be full-time and that men and women who are hired as full-time employees are equally likely to have two or more WRG records.²²²

Finally, Dr. Ward contends that his suggested method of verifying coder reliability by using match rates is in fact supported by I/O psychology research.²²³

Human Resources

The opinions of Dr. Lundquist and Dr. Outtz regarding the inadequacies of Sterling's Human Resources department are addressed by Sterling's expert, Dr. Stockdale.

Dr. Stockdale reviewed Sterling's policies, complaint investigation procedures and training programs and concluded that they are "consistent with advice provided by scholars and expert agencies such as the EEOC and OCR [U.S. Department of Education Office of Civil Rights]." ²²⁴ Dr. Stockdale asserts that Dr. Outtz "dismisses or ignores evidence that Sterling engages in strong human resource practices to curtail and to effectively respond to complaints of discrimination and harassment," including its training program for managers.²²⁵ Dr. Stockdale reviewed 46 specific complaints of unwanted sex-related behavior, as well as the workload of Regional HR Specialists and concluded that "Sterling devotes adequate resources to manage

²²¹ *Id.*

²²² *Id.* ¶5.

²²³ *Id.* ¶8 (citing authority).

²²⁴ Stockdale Report at 9, 53-60.

²²⁵ *Id.* at 8, 59.

complaints of unwanted sex-related behavior,” and that Sterling’s HR employees follow prescribed procedures in conducting investigations.²²⁶

Dr. Stockdale’s report with respect to HR is addressed in Dr. Outtz’s rebuttal report.

With respect to Sterling’s policies prohibiting sexual harassment and fraternization between managers and subordinate employees, Dr. Outtz asserts that “the mere existence of such policies does not in and of itself diminish the likelihood of Unwanted Sex-Related Behavior,” noting that Dr. Stockdale “points out in her published research that zero tolerance policies come across as tough sounding but may mask an inability to get to the root cause of sexual harassment or promote healthy workplaces,” but that she made no attempt to determine the effectiveness of the procedures used by Sterling.²²⁷ He contends, moreover, that the presence of such policies at Sterling “has not been shown to have had an appreciable effect on curbing Unwanted Sex-Related Behavior or its effects on other employees in the workplace.”²²⁸

Dr. Outtz notes that Dr. Stockdale’s opinion that Sterling’s complaint investigation process is “systematic and comprehensive” is based on a review of a sample of sexual harassment investigations selected and provided by Sterling’s counsel, that Dr. Stockdale had “no idea how many complaints were investigated at Sterling, or how representative [the sample was] of the universe of employee complaints,” that Dr. Stockdale reviewed none of Claimants’ declarations,²²⁹ and that Dr. Stockdale testified she had made no attempt to determine the effectiveness of Sterling’s investigative procedures.²³⁰

²²⁶ *Id.* at 9, 53-64 and Table E-1. Dr. Stockdale asserts that Dr. Outtz “follows no known scientific practice to arrive at his conclusion” that Sterling devotes inadequate resources to handling complaints of unwanted sex-based behavior and other “Level 3” complaints, which Dr. Outtz based on the “sheer volume of calls coming in and going out of that office.” *Id.* 61, 62-63.

²²⁷ Outtz Rebuttal Report at 7-8.

²²⁸ *Id.* at 7.

²²⁹ *Id.* at 12.

²³⁰ *Id.* at 8.

Gender-Negative Culture

The expert evidence of gender-negative culture presented by Dr. Outtz is addressed primarily by Dr. Stockdale. Sterling's critique also relies upon Dr. Outtz's deposition testimony.

Dr. Stockdale contends that Dr. Outtz's opinion that Sterling maintains an organizational culture that demeans and devalues women is not based on an accepted scientific method for measuring such a culture or climate.²³¹ Dr. Stockdale asserts that "[t]he scientifically appropriate method for assessing the psychological climate of a work environment is to use validated survey measures and sound survey research methods for collecting data from employees on their climate perceptions," such as the Organizational Tolerance for Sexual Harassment Inventory.²³² In particular, Dr. Stockdale states that the process by which Dr. Outtz arrived at his conclusion that "notwithstanding the presence of women in management and policies prohibiting sex discrimination and fraternization, an organizational climate, such as that at Sterling, facilitates Unwanted Sex-Related Behavior"²³³ "does not represent any scientific process that has ever been published in the scientific literature, presented at professional conferences, or used by practitioners of which I am aware."²³⁴

Dr. Stockdale specifically criticizes Dr. Outtz's reliance on Claimants' declarations²³⁵ and his failure to consider employee opinion surveys conducted by Sterling from 2004 through 2008; based upon Dr. Stockdale's analysis, the quantitative and qualitative data from these surveys "provides no evidence to substantiate a claim that Sterling perpetuates a climate hostile

²³¹ Stockdale Report at 8, 42, 45-49.

²³² *Id.* at 45-47.

²³³ Outtz Report at 29.

²³⁴ Stockdale Report at 52.

²³⁵ *Id.* at 49 (Dr. Outtz relied on selective declarations prepared by Complainants' attorneys of putative class members; he has "taken the worst examples of workplace mistreatment and generalized this to the whole company without any scientific basis for doing so").

or demeaning towards women.”²³⁶

Dr. Stockdale faults Dr. Outtz’s analysis of role modeling for failing to credit the positive role modeling emphasis communicated by Sterling executives through training material and other corporate communications.²³⁷ She further asserts that Dr. Outtz has misrepresented the literature on the “trickle-down” theory of modelling because the studies relied upon by Dr. Outtz are confined to supervisors and their immediate subordinates.²³⁸

Finally, Dr. Stockdale maintains that Dr. Outtz’s conclusion that a gender-negative cultural climate causes or results in discretionary pay and promotion decisions that adversely affect women has no basis in scientific literature.²³⁹

Sterling also argues that Dr. Outtz’s deposition testimony fails to establish a causal connection between the alleged gender-negative corporate culture and the pay and promotion decisions at issue in this case.

Specifically, Sterling cites the following testimony from Dr. Outtz’s deposition:

Q. So you cannot answer the question and you were not asked to answer the question of whether the supposed culture at Sterling affected 100 percent, 50 percent, 30 percent or 10 percent of the starting pay decisions at Sterling?

A. I was not asked to put a number on that and I did not address putting a number on that. What I have testified to is by virtue of this policy that was company-wide, I concluded that it was a factor in wherever these decision were made and, as you say, there were – how many thousands of them did you say were made?

Q. Thousands and thousands.

A. Within those thousands and thousands, if they considered that policy and used that factor which disadvantaged women, then it influenced their decision, and since they had discretion could have caused them to give lower pay to women.

Q. And the same with promotions, you cannot answer and you were not asked to answer whether this culture that you found at Sterling affected 5 percent, 50 percent, 90 percent, 100 percent of the promotions made during the class period?

²³⁶ *Id.* at 8, 50-52.

²³⁷ *Id.* at 8, 43-44.

²³⁸ *Id.* at 8, 44-45.

²³⁹ *Id.* at 9, 65-66.

A. I wasn't asked that question.²⁴⁰

Dr. Outtz testified that he was not asked to address how many male or female managers at Sterling were affected by the culture he found at Sterling,²⁴¹ and testified that the culture at Sterling can, but also may not influence compensation decisions.²⁴²

Dr. Outtz further testified:

A. You keep asking me about quantifying things. The kinds of phenomenon [sic] that are at issue and I was asked an opinion about in this case typically are not quantifiable, nor is it required to quantify them as to the exact number of people who did something. That's not knowable. At the end of the day one can measure—I wasn't asked to do this, but one can measure the effect of whatever this phenomenon is on differences in compensation or pay or whatever. That becomes the question where there is some quantity. But in my opinion quantifying how many of these managers actually thought about this or how many of them actually intended to do so or which ones intended to do so based on this phenomenon is not something that is quantifiable typically in my profession, no attempt would be made to quantify it nor would it be relevant to try to quantify it.²⁴³

Finally, Dr. Outtz testified:

Q. So, in other words, someone has to tell me what's causing supposed disparities and you are not the guy who can tell me that, right?

A. I'm not the guy who can tell you anything, number one. And number two, what I have testified to is that someone would look at the ultimate effect of this in terms of differences by gender. That would typically be a statistician. In my profession in the social sciences for any phenomenon, even in the research, the researcher is not going to then parse out which individuals in their minds did certain things. That's not really knowable, nor is it relevant, needed to be known.²⁴⁴

In his rebuttal report, Dr. Outtz reiterates his opinion that the evidence in the record of behavior and comments about women and women employees attributed to executives and senior managers at Sterling is "sufficient to have established workplace norms for guiding the behavior

²⁴⁰ Deposition of Dr. James Laurence Outtz (August 28, 2013) (Exhibit 1 to Sterling's Outtz *Daubert* Motion) (Outtz Dep.) at 57:24 – 159:3.

²⁴¹ *Id.* at 159:10-21.

²⁴² *Id.* at 229:12-15.

²⁴³ *Id.* at 266:15 – 267:10.

²⁴⁴ *Id.* at 268:10-24.

of managers elsewhere in the organizational hierarchy.”²⁴⁵ He points out that Dr. Stockdale has published an article on sexual harassment that acknowledges the influence of behavior by leaders of an organization on those in lower positions, and that she acknowledged in her deposition that leaders are role models whose actions have a guiding influence on members of an organization.²⁴⁶ Dr. Outtz notes that Dr. Stockdale’s evidence of positive behavior consists solely of training materials and written policies--which she herself has concluded do not, by themselves, have an appreciable effect on curbing unwanted sex-related behavior-- and that in forming her opinions, Dr. Stockdale failed to interview a single Sterling manager or employee, and did not review any of the Claimants’ declarations.²⁴⁷ Dr. Outtz also asserts that the employee surveys analyzed by Dr. Stockdale “provided no information about sexual harassment and were not formulated to elicit the kind of information normally sought in instruments accepted as appropriate measures of identifying workplace tolerance for sexual harassment.”²⁴⁸

With respect to Dr. Stockdale’s critique of his methodology, Dr. Outtz states that the examination of the declarations provided by Claimants and by Sterling, the deposition transcripts of Sterling upper management, review of Sterling’s policies and procedures, and application of relevant literature in I/O psychology to the evidence “are methods that I commonly and have seen used by I/O psychologists who have served as expert witnesses in the cases in which I have been involved over the past 30 years.”²⁴⁹

Finally, Dr. Outtz concludes that “[i]t is more likely than not that the Unwanted Sex Related Behavior that demeans and devalues women is modeled from Sterling’s executives

²⁴⁵ Rebuttal Report at 3-4.

²⁴⁶ *Id.* at 2-3.

²⁴⁷ *Id.* at 4-5.

²⁴⁸ *Id.* at 13.

²⁴⁹ *Id.* at 12. Dr. Outtz does not reference any professional literature supporting his methodology or provide citations to cases in which his methodology has been accepted.

down, influenced pay and promotion decisions to women's detriment at Sterling and would be consistent with any gender disparities in pay and promotion at Sterling."²⁵⁰

DAUBERT MOTIONS

Rule 702 of the Federal Rules of Evidence governs admission of expert testimony and provides:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

The admissibility of expert evidence is governed by the Supreme Court's decision in *Daubert*,²⁵¹ which requires that trial courts conduct a "rigorous" analysis of both the relevance and reliability of proposed expert testimony before admitting or considering such testimony. In *Wal-Mart Stores, Inc. v. Dukes* (*Wal-Mart*), the Supreme Court suggested that a *Daubert* analysis of proffered expert testimony is appropriate at the class certification stage,²⁵² and since that decision, courts in the Second Circuit have concluded that a *Daubert* analysis is required.²⁵³

Daubert requires the court to make "a preliminary assessment of whether the reasoning or methodology underlying the [expert's] testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue."²⁵⁴ The *Daubert* inquiry has been described as "flexible," and *Daubert* gives the district court the discretion needed to ensure that the courtroom door remains closed to junk science while admitting reliable expert

²⁵⁰ *Id.* at 9.

²⁵¹ 509 US 579 (1993); see also *Kumho Tire Co. v. Carmichael*, 526 US 137 (1999).

²⁵² 131 S.Ct. 2541, 2553-54 (2011).

²⁵³ See, e.g., *Lujan v. Cabana Mgmt., Inc.*, 284 F.R.D. 50, 64 (E.D.N.Y. 2012); *Floyd v. City of New York*, 283 F.R.D. 153, 166-67 (S.D.N.Y. 2012).

²⁵⁴ *Daubert*, 509 U.S. at 592-93.

testimony that will assist the trier of fact.²⁵⁵ When considering motions to exclude expert testimony, courts must bear in mind that “[v]igorous cross examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky, but admissible evidence,” not exclusion of expert testimony via *Daubert* motions.²⁵⁶ In view of liberal thrust of Federal Rules of Evidence and the presumption of admissibility of expert testimony, doubts about usefulness of expert testimony should be resolved in favor of admissibility,²⁵⁷ and only serious flaws in reasoning or methodology will warrant exclusion.²⁵⁸ Generally, “[d]isputes as to the strength of [an expert’s] credentials, faults in his use of [a particular] methodology, or lack of textual authority for his opinion, go to the weight, not the admissibility of his testimony.”²⁵⁹ At the same time, however, “[a] court may conclude that there is simply too great an analytical gap between the data and the opinion proffered,”²⁶⁰ and “when an expert opinion is based on data, a methodology, or studies that are simply inadequate to support the conclusions reached, *Daubert* and Rule 702 mandate the exclusion of that unreliable opinion testimony.”²⁶¹ Where the judge is the trier of fact, a court has greater flexibility in satisfying its gatekeeping function.²⁶²

Dr. Lanier

Sterling does not object to the acceptance of Dr. Lanier as a qualified expert in labor economics. Sterling’s *Daubert* motion to exclude Dr. Lanier’s reports and testimony is based primarily on its contentions that Dr. Lanier’s report incorrectly applied his statistical

²⁵⁵ *Amorgianos v. Nat’l R.R. Passenger Corp.*, 303 F.3d 256 (2d Cir. 2002).

²⁵⁶ *Daubert* 509 U.S. at 596.

²⁵⁷ *Canino v. H.R.P., Inc.*, 105 F. Supp. 2d 21, 28 (N.D.N.Y. 2000).

²⁵⁸ *EEOC v. Bloomberg L.P.* 2010 U.S. Dist. LEXIS 92511 at *17 (S.D.N.Y. August 31, 2010).

²⁵⁹ *McCulloch v. H.B. Fuller Co.*, 61 F.3d 1038, 1044 (2d Cir. 1995).

²⁶⁰ *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997).

²⁶¹ *Amorgianos*, 303 F.3d at 266.

²⁶² See, e.g., *Bank of N.Y. Mellon Trust Co. v. Solstice ABS CBO II, Ltd.*, 910 F.Supp.2d 629, 639 (S.D.N.Y. 2012).

methodology, incorporated false factual assumptions, and relied on an incomplete and inaccurate database created by Claimants' expert Dr. Lundquist.²⁶³

Sterling first faults Dr. Lanier's analysis because it "fails to control for significant non-discriminatory factors that account for the purported disparity he observed, namely applicants' sales productivity in their prior jobs, or, in the case of a manager, the sales volume of the store they managed."²⁶⁴

Second, Sterling faults Dr. Lanier's analysis because he relied on Dr. Lundquist's "seriously flawed" database (addressed in Sterling's motion to exclude the report and testimony of Dr. Lundquist), in particular Dr. Lundquist's failure to consider prior sales volume.²⁶⁵

Third, Sterling contends that Dr. Lanier's analysis of starting pay and first year sales should be excluded because of its flawed methodology. Sterling asserts that Dr. Lanier "erroneously ran the regression in reverse, thus arriving at meaningless and unsupported conclusions."²⁶⁶

Fourth, Sterling contends that Dr. Lanier's promotion analyses are unreliable, essentially because he excludes a majority of promotions that took place at Sterling during the relevant time period, based upon concerns regarding the validity of CAR, as well as other asserted methodological errors that are detailed in Dr. Ward's report—in which Dr. Ward concludes (based upon CAR data) that women interested in and available for promotion have been promoted at the same rate as men.²⁶⁷

²⁶³ Sterling Jewelers Inc.'s Motion to Exclude the Report and Testimony of Dr. Louis Lanier (Sterling's Lanier *Daubert* Motion) at 1.

²⁶⁴ *Id.* at 8.

²⁶⁵ *Id.* at 12.

²⁶⁶ *Id.* at 14.

²⁶⁷ *Id.* at 19-21.

Fifth, Sterling contends that Dr. Lanier's "doubts" regarding CAR are unfounded and not grounded in any scientific analysis.²⁶⁸ Sterling notes that the source of Dr. Lanier's concern is reflected in his Table 13, which shows that 42.3 percent of promotions took place within one month of the candidate registering in CAR, which suggested to Dr. Lanier that promoted individuals might have been selected for management and then instructed to register in CAR, and thus that CAR does not reveal true interest in promotion. Sterling points out that Dr. Lanier did not examine the effect of this phenomenon on women, but that an analysis conducted by Dr. Ward shows that there were significantly more women than men who were promoted within short time periods after registration.²⁶⁹

Sixth, Sterling contends that Dr. Lanier's merit pay analyses are incorrect and rely on faulty assumptions.²⁷⁰ Sterling contends that Dr. Lanier incorrectly calculated average merit pay, and that, properly calculated, women sales associates receive higher merit pay as compared to men, which is consistent with Dr. Lanier's finding that women receive, on average, higher performance ratings. Sterling further contends that Dr. Lanier's analyses failed to account for an important characteristic of Sterling's merit pay system, the "percent to pay" benchmark, which acts as a cap on pay for higher paid sales associates even if they meet their goal.²⁷¹

Finally, Sterling asserts that Dr. Lanier's Table 15, which analyzes base pay and promotions by region and by district, is irrelevant because it fails to control for any significant non-discriminatory variables, including the prior experience in the Lundquist database.²⁷²

²⁶⁸ *Id.* at 21-23.

²⁶⁹ *Id.* at 23.

²⁷⁰ *Id.* at 24.

²⁷¹ *Id.* at 25.

²⁷² *Id.* at 26.

Claimants oppose Sterling's motion, relying in part on the rebuttal report of Dr. Lanier, which includes the results of additional statistical studies conducted by him that are described above.

With respect to Dr. Lanier's failure to consider prior sales volume, Claimants note that there is no prior sales volume data for 2003 to 2008, and that there is only prior sales volume data for 2009-2012 for about 29% of employees, and that the reliability of WRG inputs is questionable.²⁷³ In any event, Dr. Lanier analyzed the relationship between the WRG generated wage rates and actual sales volume (Table 9), finding that women were paid less than men who had the same level of sales, or the same ratio of sales to sales goal, indicating that the WRG is a poor predictor of productivity.²⁷⁴ Claimants further contend that Dr. Lanier properly relied on Dr. Lundquist's applicant study, for which Dr. Lundquist used appropriate coding methodology and which permitted Dr. Lanier to consider separately the impact of prior jewelry sales experience and prior non-jewelry sales experience. Claimants assert that Sterling's characterization of Dr. Lanier's analysis in Table 9 as a "reverse regression" is "simply incorrect," and that it is in fact Dr. Ward who conducted a reverse regression.²⁷⁵ With respect to Dr. Lanier's promotion analysis, Claimants argue that Dr. Lanier properly excluded individuals with missing evaluation data to ensure he was comparing similarly situated individuals.²⁷⁶ Nonetheless, Dr. Lanier re-ran his promotion analyses including the omitted individuals and obtained results similar to his original report.²⁷⁷ With respect to CAR, Claimants note that Dr. Lanier reasonably considered evidence that cast doubt on the reliability of CAR as an accurate

²⁷³ Opposition to Motion to Strike Expert Report of Dr. Louis Lanier (Claimants' Lanier *Daubert* Response) at 5-8.

²⁷⁴ *Id.* at 8.

²⁷⁵ Claimants' Lanier *Daubert* Response at 9-11.

²⁷⁶ *Id.* at 12.

²⁷⁷ *Id.* at 13.

measure of interest in promotion, *e.g.*, patterns of registration suggesting that employees were first selected for promotion and then told to register in CAR to make the promotion permissible.²⁷⁸ With respect to Dr. Lanier's merit pay analysis, Claimants note that although Drs. Lanier and Ward used a different methodology to exclude from consideration individuals who were not eligible for a merit increase, they reached the same conclusion: there are no statistically significant differences with respect to merit increases, if performance evaluations are considered.²⁷⁹ Finally, Claimants contend that Sterling has not shown that correction of any of the other purported errors in Dr. Lanier's report would yield different results.²⁸⁰

In reply, Sterling argues that Dr. Lanier's rebuttal report concedes that his initial starting pay analysis failed to control for a significant non-discriminatory factor affecting pay (personal jewelry sales volume), and that by using WRG data in his rebuttal analysis, Dr. Lanier has acknowledged the insufficiency of the data used in his initial analysis.²⁸¹ Sterling argues that Dr. Lanier's revised starting pay analysis is unreliable because he disregarded all of the WRG components other than personal jewelry sales volume, in particular prior managerial experience, which Dr. Ward shows is highly correlated and predictive of future sales.²⁸² Sterling continues to maintain that Dr. Lanier's analysis of starting pay and sales (Lanier Report, Table 9) uses a flawed methodology leading to unreliable results.²⁸³

²⁷⁸ *Id.* at 13-14.

²⁷⁹ Claimants also state that Dr. Lanier re-ran his primary analyses, taking Dr. Ward's critiques into account and reached the same conclusions. Claimants' Lanier *Daubert* Response at 15-16. However, Sterling points out that there is no revised version of that analysis in Dr. Lanier's Rebuttal Report. Sterling Jewelers Inc.'s Reply Memorandum in Support of Its Motion to Exclude the Reports and Testimony of Dr. Louis Lanier (Sterling's Lanier *Daubert* Reply) at 13, n.44.

²⁸⁰ Claimants' Lanier *Daubert* Response at 16-17.

²⁸¹ Sterling's Lanier *Daubert* Reply at 18-19.

²⁸² *Id.* at 7-8.

²⁸³ *Id.* at 9-11.

With respect to merit pay, Sterling contends that its criticisms of Dr. Lanier's initial report "stand un rebutted," and that Dr. Lanier's rebuttal analysis, which concludes that merit pay increases do not result in equal overall pay for equally productive men and women, confuses employee performance ratings with employee sales.²⁸⁴

With respect to promotions, Sterling asserts that Dr. Lanier has abandoned his initial promotions analysis thereby conceding that the analysis lacked probative value, and that Dr. Lanier fails to identify any scientific basis for his "doubts" regarding CAR.²⁸⁵ Sterling contends that there is no evidence that women are discouraged from expressing their interest in promotion through CAR, and that the fact that women comprise 73 percent of Sales Associates and 72 percent of Assistant Managers "belies the notion that women are somehow discouraged from promotion."²⁸⁶

Applying the general principles governing *Daubert* motions set forth above, I find that none of the deficiencies asserted by Sterling requires the exclusion of Dr. Lanier's reports or testimony. Reliability "is primarily a question of the validity of the methodology employed by an expert, not the quality of the data used in applying the methodology or the conclusions produced," and statisticians are afforded substantial latitude in employing regression analysis, including the choice of variable to include.²⁸⁷ I find that each of Sterling's criticisms goes to the probative value rather than admissibility of Dr. Lanier's report and testimony, and that Sterling's critique is best addressed through cross-examination rather than exclusion. Sterling's motion to exclude the reports and testimony of Dr. Lanier is therefore denied.

²⁸⁴ *Id.* at 14-15.

²⁸⁵ *Id.* at 19-20.

²⁸⁶ *Id.* at 20.

²⁸⁷ *Manpower, Inc. v. Insurance Company of Pennsylvania*, 732 F.3d 796, 806-809 (7th Cir. 2013).

Dr. Lundquist

Sterling does not object to the acceptance of Dr. Lundquist as a qualified expert in Industrial/Organizational Psychology and Psychometrics. Sterling's *Daubert* motion to exclude Dr. Lundquist's reports and testimony faults her methodology and conclusions on multiple grounds, many of which are set forth above.

With respect to Dr. Lundquist's initial report, Sterling first contends that Dr. Lundquist's coding study is fundamentally flawed and unreliable. At the outset, Sterling faults Dr. Lundquist's failure to use WRG data (available after July 2009), as opposed to job applications, which frequently contain ambiguous or vague information with respect to prior experience. Sterling contends that Dr. Lundquist's coders had to use default rules and their personal judgments about the applicant's prior experience, and that the WRG data is more reliable because DMs had the opportunity to clarify such information in an interview or other follow-up. Sterling also faults Dr. Lundquist's failure to measure or test the accuracy of the coding against actual known results (*e.g.*, by interviewing a representative sample of applicants), and asserts that the methodology used by Dr. Lundquist to test consistency of coding among the coders was highly flawed and resulted in a misleading assessment of coder reliability (providing examples of inaccuracies and inconsistent coding by Dr. Lundquist's team).²⁸⁸ Sterling further faults Dr. Lundquist's failure to account for prior sales volume (available from the WRG data, but not job applications), which she concedes is a reasonable factor for Sterling to consider, and which Dr. Ward determined is statistically the most important determinative factor for starting base pay.²⁸⁹

²⁸⁸ Sterling's Motion to Exclude the Report and Testimony of Dr. Kathleen Lundquist (Sterling's Lundquist *Daubert* Motion) at 18-25. Sterling cites several cases in which courts have excluded expert testimony based upon unreliable coding. *Id.* at 22-25. I find that these cases are distinguishable for substantially the reasons set forth in Claimants' Opposition to Motion to Strike Expert Report of Dr. Kathleen Lundquist (Claimants' Lundquist *Daubert* Response) at 27-28.

²⁸⁹ Sterling's Lundquist *Daubert* Motion at 13. Dr. Lundquist did not consider prior sales volume in her

Sterling next contends that Dr. Lundquist's opinions about the job-relatedness of various prior experience factors, particularly prior management experience, are unreliable. Sterling faults Dr. Lundquist's failure to perform a job analysis, her failure to analyze all of the relevant and available O*NET data, her failure to perform any empirical or statistical study to determine job-relatedness, and the inclusion of "sales support" as a job-related factor, which Sterling argues includes important elements of managerial experience. Sterling also points out that Dr. Lundquist's conclusions are contradicted by Dr. Ward's statistical analyses, which show that prior managerial experience is highly correlated with sales at a statistically significant level, and that Dr. Lundquist's customer service fields are not related to the sales productivity of Sterling employees, and that the five-year experience cap is not adverse to women.²⁹⁰

In rebuttal, Dr. Lundquist asserts that the methods and principles used to code applications in her applicant study "are regularly used by Industrial/Organizational Psychologists to evaluate applicants' training and experience in both research and practice."²⁹¹ Dr. Lundquist provides a detailed explanation of her approach to testing inter-coder consistency, as well as a critique of Dr. Ward's suggested approach,²⁹² noting that Dr. Ward is concededly not an expert on coding, and that Sterling's expert Dr. Stockdale used the same technique in her report to

initial report because she concluded that prior sales volume was not consistently documented and verified by Sterling's DMs. Dr. Ward notes that Dr. Lundquist did not analyze the WRG data to test this opinion. Dr. Ward analyzed 7,000 applications reporting prior sales volume and determined that this factor was consistently verified by Sterling's managers and that there is ample documentation in the WRG database to confirm the legitimacy of this verification process. *Id.* at 13, n.47.

²⁹⁰ *Id.* at 29. Sterling also argues that Dr. Lundquist's opinions improperly encroach on factual determinations reserved for the arbitrator and lack a scientific basis. Specifically, Sterling contends that Dr. Lundquist makes improper lay findings of fact and credibility determinations regarding Sterling's HR Department without any purported scientific basis, and that Dr. Lundquist makes improper lay findings of fact and unreasonable extrapolations regarding Sterling's Career Advancement Registry. This argument is rejected for substantially the reasons set forth in Claimants' Lundquist *Daubert* Response at 29-34.

²⁹¹ Lundquist Rebuttal Report at 31 (citing authorities).

²⁹² *Id.* at 31-34.

assess the reliability of the coding of employee survey comments.²⁹³

With respect to Sterling's critique of Dr. Lundquist's failure to use WRG data, Dr. Lundquist conducted several additional studies, using a hierarchical multiple regression, in which she analyzes the same WRG data used by Dr. Ward (including prior experience and prior sales volume).²⁹⁴ Dr. Lundquist also controlled for store volume, which she contends that Dr. Ward improperly failed to consider. She concludes in her rebuttal report that "other management" and "store management" are not predictive of annual sales and should not have been used in the WRG. With respect to prior sales volume, she concludes that only "personal sales jewelry volume" and "personal non-jewelry sales volume" had statistically significant relationships with the performance of full-time Sales Associates and that the variable with the most predictive power was "personal sales jewelry volume."²⁹⁵ Dr. Lundquist found significant gender differences in variables that are not predictive of job performance, which would cause compensation to be affected differentially for men and women on the basis of invalid experience factors.²⁹⁶

In reply, Sterling identifies numerous asserted flaws in Dr. Lundquist's analysis of WRG data. Sterling asserts that (1) in order to show that the correlation of managerial experience variables with future sales is not statistically significant, Dr. Lundquist "moves the goal post" as to what counts for statistical significance;²⁹⁷ (2) Dr. Lundquist's use of store volume as a control

²⁹³ *Id.* at 40-41.

²⁹⁴ Dr. Lundquist continues to contend that the WRG data are not reliable due to inconsistent inputs and because the DMs' failure to follow WRG procedures, which compromised the accuracy and job-relatedness of the company's starting pay decisions. Lundquist Rebuttal Report at 3-4; 23-30.

²⁹⁵ *Id.* at 17-20.

²⁹⁶ *Id.* at 21.

²⁹⁷ Sterling Jewelers Inc.'s Reply to Claimants' Response to Sterling's Motion to Exclude the Report and Testimony of Dr. Kathleen Lundquist (Sterling's Lundquist *Daubert* Reply) at 11. In her initial report, Dr. Lundquist used two standard deviations as the standard for statistical significance. "Social scientists consider a finding of two standard deviations significant, meaning there is about one chance in 20 [5%]

measure violates the basic tenets of statistical regression analysis; (3) Dr. Lundquist's use of "hierarchical regression" is misleading because it produces different results depending on the order of variables; and (4) Dr. Lundquist provides no valid rationale for examining full-time and part-time employees separately.²⁹⁸

Overall, Sterling raises a number of significant questions regarding Dr. Lundquist's approach and conclusions that may well diminish the probative value of Dr. Lundquist's testimony. I find, however, that Sterling's criticism of Dr. Lundquist's application of recognized methodologies (application coding and regression analysis) clearly goes to the weight to be given to Dr. Lundquist's opinions, as opposed to their admissibility. *McCulloch*, 61 F.3d at 1044.

Dr. Outtz

Sterling contends that Dr. Outtz's report and testimony should be excluded because (1) his opinions and conclusions cannot assist the trier of fact; (2) he is not qualified to offer the types of opinions and testimony he has put forth; and (3) his analysis is "wholly unreliable."²⁹⁹

Sterling asserts that Dr. Outtz's opinions and conclusions cannot assist the trier of fact because he cannot testify that the modeling he describes impacted any pay or promotion decision at Sterling, and because he cannot establish a link between alleged sexual harassment and pay or

that the explanation for a deviation could be random and the deviation must be accounted for by some factor other than chance." *Waisome v. Port Auth.*, 948 F.2d 1370, 1376 (2d Cir. 1991) (citations omitted). In her rebuttal report, Dr. Lundquist uses a standard deviation value corresponding to the likelihood that the same or greater differential would occur by chance less than 1% of the time (1 out of 100) with respect to her conclusion that managerial experience does not predict future sales. Sterling's Lundquist *Daubert* Reply at 11. Dr. Lundquist's rationale for using a different measure was that the sample size in this case was very large. Lundquist Rebuttal Report at 21, n.11; Sterling's Lundquist *Daubert* Reply, Exhibit 4 (Lundquist Dep. II at 153:15-23).

²⁹⁸ Sterling's Lundquist *Daubert* Reply at 15. Sterling also faults Dr. Lundquist's assessment of the reliability or consistency of WRG inputs, which Sterling contends is based on misunderstandings of company policy and speculation and cannot assist the arbitrator because Dr. Lundquist fails to identify any deviation from procedure or inconsistency that disadvantaged female applicants. *Id.* at 16-20. This critique raises some issues that may well affect the probative value of Dr. Lundquist's conclusions, but which do not warrant exclusion of her testimony.

²⁹⁹ Sterling Jewelers Inc.'s Motion to Exclude the Report and Testimony of Dr. James Outtz (Sterling's Outtz *Daubert* Motion) at 2.

promotion decisions. Specifically, Sterling argues that Dr. Outtz's inability to quantify the impact of the allegedly gender-negative culture at Sterling requires exclusion of his report and testimony based upon the ruling in *Wal-Mart* with respect to the testimony of sociologist Dr. William Bielby, which the Court found was not sufficient, by itself, to establish commonality, because Dr. Bielby could not say what percentage of employment decisions was determined by stereotyped thinking.³⁰⁰ Here, however, Claimants do not rely solely on the testimony of Dr. Outtz. Moreover, following *Wal-Mart*, similar testimony was found to be relevant in *Ellis v. Costco*, by providing support for the claim that Costco operated under a common, companywide promotion system.³⁰¹ I therefore find that Dr. Outtz's testimony may be of assistance to the trier of fact, and that his inability to link the allegedly gender-negative culture to specific pay or promotion decisions goes to the probative value rather than the admissibility of his opinions.

Sterling asserts that Dr. Outtz has "no expertise in the fields of gender discrimination, workplace behavior, workplace norms, modeling of workplace behavior, or the effects of workplace behavior on tangible employment decisions such as pay or promotion."³⁰² Specifically, Sterling notes that Dr. Outtz has never published any works pertaining to sexual harassment, and that prior to his retention in this case, Dr. Outtz had "never been asked to address whether manager or executive behavior established workplace norms that guided behavior elsewhere in the company."³⁰³

An expert must stay within the reasonable confines of his subject area, and cannot render expert opinion on an entirely different field or discipline.³⁰⁴ At the same time, "[l]iberality and

³⁰⁰ *Wal-Mart*, 131 S.Ct. at 2553-54.

³⁰¹ 2012 U.S. Dist. LEXIS 13718 at *89 (N.D. Ca. September 25, 2012).

³⁰² Sterling's Outtz *Daubert* Motion at 13.

³⁰³ *Id.* at 14.

³⁰⁴ *Lappe v. Am. Honda Motor Co., Inc.*, 857 F.Supp. 222, 227 (N.D.N.Y. 1994), *aff'd sub nom.*, *Lappe v. Honda Motor Co. Ltd. of Japan*, 101 F.3d 682 (2d Cir. 1996).

flexibility in evaluating qualifications should be the rule; the proposed expert should not be required to satisfy an overly narrow test of his own qualifications,” and a lack of specialization affects the weight of the opinion, not its admissibility.³⁰⁵ Dr. Outtz received his Ph.D. in Industrial/Organizational Psychology from the University of Maryland in 1976, and is a “Fellow” of the Society for Industrial and Organizational Psychology (SIOP), the American Psychological Association, and the American Educational Research Association. He has been qualified as an expert in ten cases involving gender discrimination claims and over 60 cases involving discrimination claims, including in the case of *Velez v. Novartis*, a class action gender discrimination case brought in the Southern District of New York by female sales associates involving disparate pay, failure to promote, and sexual harassment claims; in that case the court found Dr. Outtz qualified as an expert in the “field of employment practices and procedures relating to hiring, development, evaluation and promotion of employees.”³⁰⁶ Additionally, Dr. Outtz has provided expert opinions on organizational culture and modeling of behavior by executives and managers in several cases, including *McReynolds v. Merrill, Lynch, Fenner, Pierce and Smith* and *Gutierrez v. Johnson & Johnson*.³⁰⁷ Based upon these credentials and this experience, I find that Dr. Outtz’s report and testimony in this case fall within the “reasonable confines” of his expertise, and that any deficiencies in his qualifications go to the probative value and not the admissibility of his expert opinion.

Finally, Sterling asserts that Dr. Outtz’s opinions and testimony should be excluded because they are unreliable. Sterling contends that Dr. Outtz drew unsupported inferences from insufficient sampling, failed to apply any scientific methodology to his analysis, and ignored

³⁰⁵ *Id.*; *Tiffany (NJ) Inc. v. eBay, Inc.*, 576 F.Supp.2d 457, 458-458 (law does not require a narrow specialty; objections to qualifications go to the weight not the admissibility of expert’s testimony, and are more properly explored on cross-examination).

³⁰⁶ *Velez v. Novartis*, 2010 US Dist. LEXIS 95010 at *19 (S.D.N.Y. February 25, 2010).

³⁰⁷ Outtz Report at Appendix 2; Outtz Rebuttal Report at 11.

non-conforming evidence that contradicted his theory, and that his modeling theory is not supported by any scientific literature.³⁰⁸

Sterling notes that the declarations of 193 female declarants that form the cornerstone of Dr. Outtz's opinions and conclusions represent 0.44 percent of the entire putative class of 44,000 women. In addition, Sterling asserts that Dr. Outtz did not apply any recognized methodology (scientific or otherwise) to his examination of the declarations—all of which he found to be credible, without using any methodology to verify their accuracy. Moreover, in selecting the ten declarants that he interviewed, Dr. Outtz picked those who presented the strongest evidence in favor of Claimants.³⁰⁹

Sterling further contends that Dr. Outtz's testimony and report are unreliable because he ignored evidence in the record that failed to fit his theory. As noted above, Dr. Outtz discounted the relevance of a 2006 third-party employee satisfaction survey because it did not specifically address sexual harassment. Sterling contends, however, that the survey reflects positive attitudes on items that would have captured concerns about a hostile or demeaning work environment conducted by a third-party vendor. Sterling also faults Dr. Outtz for not requesting other employee satisfaction surveys. Sterling also criticizes Dr. Outtz's failure to acknowledge the effect of women in managerial positions on his conclusion that managers model unwanted sex-related behavior at higher managerial levels, given that 43.2 percent of DMs and 32.4 percent of VPROs were women, and noting that Dr. Outtz conceded at his deposition that women would be less likely than men to model unwanted sex-related behavior.³¹⁰

³⁰⁸ Sterling's Outtz *Daubert* Motion at 14-15.

³⁰⁹ *Id.* at 15-19.

³¹⁰ *Id.* at 19-23.

Sterling also argues the Dr. Outtz drew unsupported inferences regarding Sterling's HR function and failed to apply any scientific methodology to his analysis. With respect to Dr. Outtz's conclusion that Sterling was unresponsive to sexual harassment complaints, Sterling notes that Dr. Outtz relied on nine declarations out of a putative class of 44,000 women, and admitted that he did not review any of the documents produced in this matter that reflect Sterling's investigations of complaints or the discipline that occurred following those investigations.³¹¹

In sum, Sterling contends that Dr. Outtz's conclusions are based on a flawed analysis because no survey was conducted, no random sampling of the Sterling employee population was performed to assess their perceptions of the organizational climate, and no systematic sampling of company records was conducted to examine how the company handles sexual harassment grievances, conducts training or otherwise monitors compliance with its policies on sexual harassment and discrimination.³¹²

Sterling further asserts that the scientific literature on which Dr. Outtz relies does not support his theory of modeling. Sterling notes that critical to Dr. Outtz's theory is his contention that alleged unwanted sex-related behavior occurs at the highest levels of the company and permeates down through multiple levels in the organization throughout the relevant period. Sterling points out that the studies to which Dr. Outtz cites focus exclusively on the relationships between supervisors and subordinates who deal directly with each other, and that none of the studies concludes that the behavior of leaders influenced the climate toward an entire class of employees. Moreover, Sterling asserts that none of the studies deals with unwanted sex-related behavior, and thus that no study establishes any connection between unwanted sex-related

³¹¹ *Id.* at 25-26.

³¹² *Id.* at 27.

behavior and pay and promotion decisions. According to Sterling, there is no scientific study establishing that a culture that is tolerant of sexual harassment or of gender inequity generally causes discrimination in pay or promotion.³¹³

In response to Sterling's critique, Claimants challenge Sterling's characterization of the record on numerous points. For example, Claimants deny that Dr. Outtz relied almost exclusively on Claimants' declarations in forming his opinions, pointing out that Dr. Outtz also reviewed the deposition transcripts of fourteen Sterling executives, numerous internal Sterling documents, and approximately 600 declarations of employees on behalf of Sterling.³¹⁴ Claimants argue that Sterling has provided no authority for its contention that Dr. Outtz was required to conduct any random or statistical sampling because he reviewed and analyzed all of Claimants' declarations, and point out that Dr. Outtz verified the accuracy of the declarations by assessing the extent to which the declarants provided first-hand accounts, the degree of detail contained in the declarations and by comparing the information provided with the testimony of Sterling's witnesses.³¹⁵ Claimants point out that Dr. Outtz addressed the effect of women in managerial positions at length in his initial report.³¹⁶ With respect to Sterling's complaint that Dr. Outtz failed to request additional employee satisfaction surveys, Claimants point out that the 2006 survey (which Dr. Outtz reviewed) was the only survey for which there was data on gender.³¹⁷ With respect to Dr. Outtz's failure to systematically assess Sterling's investigation of sexual harassment complaints, Claimants note that Sterling successfully moved to preclude

³¹³ *Id.* at 27-30.

³¹⁴ Response in Opposition to Sterling Jewelers Inc.'s Motion to Exclude the Report and Testimony of Dr. James Outtz (Claimants' Outtz *Daubert* Response) at 10-11.

³¹⁵ *Id.* at 11-12.

³¹⁶ *Id.* at 19.

³¹⁷ *Id.* at 18-19.

discovery of this evidence.³¹⁸ Finally, Claimants dispute Sterling's characterization of the studies relied upon by Dr. Outtz, pointing out that he in fact cites to scientific literature that shows that the prevalence of unwanted sex-related behavior, together with the climate and culture that facilitate it, derogates and devalues women, and citing cases permitting expert testimony similar to that offered by Dr. Outtz.³¹⁹ Claimants acknowledge that Dr. Outtz's opinions may in some respects be considered a "new theory," but argue that Rule 702 permits the admissibility of new theories provided they are based on reliable methodology.³²⁰

On balance, I find that Sterling has raised some significant concerns regarding Dr. Outtz's methodology and the reliability of Dr. Outtz's report and testimony. However, as Claimants' response demonstrates, certain aspects of the record, and Sterling's characterization of Dr. Outtz's opinions are sharply disputed. Under these circumstances, and applying the general principles governing *Daubert* motions set forth above, I find that the deficiencies identified by Sterling go to the probative value rather than the admissibility of Dr. Outtz's report and testimony.

For all of the above reasons, Sterling's motion to exclude the report and testimony of Dr. Outtz is denied.

Dr. Stockdale

Claimants do not challenge Dr. Stockdale's qualifications as an expert in I/O Psychology. They assert, however, that her report and testimony should be excluded because they cannot assist the trier of fact and because her opinions and conclusions are unreliable.

Claimants contend that Dr. Stockdale's opinions on gender role socialization are irrelevant because the statistics she cites in her report are not related to the demographics at

³¹⁸ *Id.* at 5.

³¹⁹ *Id.* at 20-21.

³²⁰ *Id.* at 21, n. 103.

Sterling or the positions at issue in this case (noting that the literature pertains largely to top management positions in unrelated fields).³²¹ Claimants also fault Dr. Stockdale's reliance on documents selected by Sterling's counsel without conducting any independent analyses.³²² For example, Dr. Stockdale did not interview any Sterling managers and did not investigate or attempt to determine how Sterling's policies and procedures are implemented *in practice*.³²³ Nor did Dr. Stockdale review any of Claimants' declarations.³²⁴ Claimants further fault Dr. Stockdale for simply accepting the conclusions of Sterling's other experts, including the assertion that the algorithm elements of the WRG are "job relevant," and "that gender differences in starting pay are fully accounted for by gender differences in prior experience, particularly management experience."³²⁵ Dr. Stockdale acknowledged that she did not have detailed knowledge of the wage engine, which was used for much of the class period prior to the introduction of the WRG.³²⁶ Claimants further assert that the professional literature on which Dr. Stockdale relies does not support her opinions, and that her own prior work conflicts with her critique of Dr. Outtz's modeling analysis.³²⁷

Finally, Claimants contend that Dr. Stockdale's methodologies are unreliable, citing her reliance on 46 investigation files selected by counsel, rather than a random sample of the full universe of investigation files, and noting that her conclusion that Sterling's personnel procedures were "routinely followed" is inconsistent with a Sterling memo (not reviewed by Dr. Stockdale) in which a consultant concluded that Sterling's HR employees "may have taken a less

³²¹ Claimants' Motion to Strike the Report and Testimony of Dr. Margaret Stockdale (Claimants' Stockdale *Daubert* Motion) at 3-6.

³²² *Id.* at 7.

³²³ *Id.* at 8.

³²⁴ *Id.*

³²⁵ *Id.* at 7.

³²⁶ *Id.* at 9.

³²⁷ *Id.* at 11-15

than robust approach to the internal investigation process.”³²⁸

In response, Sterling states that it “has not offered Dr. Stockdale’s report as a comprehensive study of its pay and promotion practices,” but to “identify deficiencies in the opinions and testimony of Dr. Outtz.”³²⁹ Sterling nonetheless proceeds to argue for the admissibility of all of Dr. Stockdale’s opinions. However, based upon the limited purpose for which Sterling offers Dr. Stockdale, I find that Opinions Nos. 3, 5, and 6,³³⁰ which broadly endorse Sterling’s pay and promotion practices, should be excluded. This addresses a number of Claimants’ objections. I find that Dr. Stockdale’s critique of Dr. Outtz’s modeling theory may assist the trier of fact and is reliable and therefore admissible, as is her critique of Dr. Outtz’s conclusions regarding Sterling’s HR Department. The other deficiencies in her critique of Dr. Outtz identified by Claimants, including her reliance on limited data, go to the weight rather than the admissibility of her testimony. Similarly, the asserted deficiencies in Dr. Stockdale’s application of gender socialization theory with respect to female interest in promotions go to the probative value rather than to the admissibility of her conclusions.

Applying the general principles governing *Daubert* motions set forth above, I find that the deficiencies asserted Claimants go to the probative value rather than the admissibility of Dr. Stockdale’s report and testimony, except that Opinions 3, 5 and 6 are excluded.

For all of the above reasons, Sterling’s motion to exclude the report and testimony of Dr. Stockdale is granted in part and denied in part.

³²⁸ *Id.* at 15-17 (citing Exhibit 8 to Claimants’ Stockdale *Daubert* Motion).

³²⁹ Sterling Jewelers Inc.’s Opposition to Claimants’ Motion to Strike the Report and Testimony of Dr. Margaret Stockdale (Sterling’s Stockdale *Daubert* Response) at 2.

³³⁰ Stockdale Report at 7-8.

Drs. Dunleavy and Sady

Claimants contend that the report of Drs. Dunleavy and Sady and the testimony of Dr. Dunleavy should be excluded under *Daubert*. Claimants challenge the qualifications of these proposed experts and argue that their report and testimony are unreliable.

Drs. Dunleavy and Sady each have an M.A. and Ph.D. in Industrial/Organizational Psychology. Neither has testified as an expert before.

Dr. Sady received his Ph.D. in 2012. He is an adjunct professor at University of Maryland, Baltimore County, where he teaches statistics. He has presented a number of papers on employment-related topics to various professional meetings, but appears not to have published any scholarly work in peer-reviewed journals, and has no discernible specialization. Based upon his limited experience and relative lack of expertise in the subjects of his report, I find that Dr. Sady is not sufficiently qualified to testify as an expert in this matter.

By contrast, Dr. Dunleavy received his Ph.D. in 2004 and has published extensively, with a concentration in the area of selection procedures, validation, and adverse impact. Dr. Dunleavy has also served on several committees relating to his field, including the SIOP Professional Practice Committee and the SIOP task force responsible for discussing contemporary selection practices with the EEOC. I therefore find that Dr. Dunleavy is qualified to testify with respect to job-relatedness and the reliability of selection processes, which are subjects clearly within the “reasonable confines” of his expertise.

Claimants contend that the opinions and conclusions contained in the report with respect to the WRG are inadmissible based upon Dr. Dunleavy’s lack of knowledge about the development and implementation of the WRG, and the failure to use any scientific or

independent methodology for assessing the reliability or job-relatedness of the WRG.³³¹

Claimants point out that at his deposition, Dr. Dunleavy did not know how Sterling developed the WRG, how Sterling decided on the factors to include in the WRG, how Sterling's DMs calculated years of experience, or how Sterling managers count retail sales experience.³³² Although the report claims that "productivity metrics" entered into the WRG were required to be verified and were "subject to internal audit," Dr. Dunleavy admitted in his deposition that he does not know whether Sterling ever conducted any verification or internal audits of the WRG inputs, and does not know what kind of documentation was required in order to prove prior sales volume.³³³ Claimants further note that Dr. Dunleavy concededly did not examine whether Sterling's compensation procedures were implemented in a reliable manner, and did not recall whether he reviewed an internal Sterling document (listed on Appendix A to his report) reporting that out of 65 new hires processed through the WRG, only 25 were processed properly.³³⁴ These asserted deficiencies are troubling, and may well limit the probative value of Dr. Dunleavy's opinions regarding reliability. I find, however, that they are best addressed by cross-examination as opposed to exclusion.

Claimants also fault Drs. Dunleavy and Sady for relying on Dr. Ward's analysis to support their conclusions, without independently verifying his work or conducting independent

³³¹ Claimants' Motion to Exclude the Report and Testimony of Drs. Eric Dunleavy and Kayo Sady (Claimants' Dunleavy/Sady *Daubert* Motion) at 1, 4-8. Claimants also maintain that the experts' opinions on job-relatedness of prior management experience based upon O*NET and research literature are unreliable because they are "completely divorced from the facts of this case," noting that Drs. Dunleavy and Sady did not review the job descriptions for the positions at issue in this case. Reply in Support of Claimants' Motion to Exclude the Report and Testimony of Drs. Eric Dunleavy and Kayo Sady (Claimants' Dunleavy/Sady *Daubert* Reply) at 5. The debate between the experts regarding which O*NET data are relevant is plainly a matter for cross-examination and not a basis for exclusion.

³³² Claimants' Dunleavy/Sady *Daubert* Motion at 4-5.

³³³ *Id.* at 6.

³³⁴ *Id.* at 8, 10.

studies.³³⁵ I disagree. Although Dr. Dunleavy observes that his conclusions are confirmed by Dr. Ward's studies, Dr. Dunleavy's analysis of job-relatedness and reliability is based primarily on his interpretation of O*NET data, and application of the scientific literature on HR process structure, and does not "merely parrot" Dr. Ward's conclusions.³³⁶ Nor is this a situation in which an expert has done no more than repeat information provided by a party.³³⁷

For all of the above reasons, Claimants' motion to strike the report and testimony of Dr. Dunleavy is denied; the motion to strike any testimony by Dr. Sady is granted.

CLASS CERTIFICATION ANALYSIS: TITLE VII

I turn first to whether Claimants have met the requirements for class certification with respect to their claims under Title VII.³³⁸

Claimants' Theories of Liability

As noted above, Claimants assert two theories of liability under Title VII: disparate impact and pattern and practice disparate treatment.

Disparate Impact

Disparate impact claims "involve employment practices that are facially neutral in their treatment of different groups but that in fact fall more harshly on one group than another and cannot be justified by business necessity."³³⁹ Proof of discriminatory motive or intent is not required.³⁴⁰ "[T]he necessary premise of the disparate impact approach is that some

³³⁵ *Id.* at 7.

³³⁶ *Cf. Eberli v. Cirrus Design Corp.*, 615 F.Supp.2d 1357, 1364 (S.D. Fla. 2009).

³³⁷ *Cf. Auther v. Oshkosh Corp.*, 2013 U.S. Dist. LEXIS 132600 at *9-14 (W.D.N.Y. September 16, 2013) and *Arista Records LLC v. Usenet.com, Inc.*, 608 F.Supp.2d 409, 424-426 (S.D.N.Y. 2009).

³³⁸ Because class certification of Claimants' EPA claims raises a number of distinct issues, the request for class certification of EPA claims is discussed separately at pp. 113-115.

³³⁹ *Int'l Bhd. of Teamsters v. United States (Teamsters)*, 431 U.S. 324, 335 n.15 (1977).

³⁴⁰ *Id.*

employment practices, adopted without a deliberately discriminatory motive may in operation be functionally equivalent to intentional discrimination”³⁴¹ “[S]ubjective or discretionary employment practices may be analyzed under the disparate impact approach in appropriate cases.”³⁴²

The following description of the proof typically presented in disparate impact cases is closely adapted from the decision of the Second Circuit in *Robinson v. Metro-North Commuter R.R.*³⁴³

Disparate impact claims involve three stages of proof. The first is the *prima facie* showing of disparate impact. It requires plaintiffs to establish by a preponderance of the evidence that the employer uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin. To make this showing, a plaintiff must (1) identify a policy or practice, (2) demonstrate that a disparity exists, and (3) establish a causal relationship between the two. Allegations solely of a bottom line racial imbalance in the work force are insufficient.

Statistical proof almost always occupies center stage in a *prima facie* showing of a disparate impact claim. Statistical proof can alone make out a *prima facie* case. The statistics must reveal that the disparity is substantial or significant and the statistics must be of a kind and

³⁴¹ *Watson v. Fort Worth Bank & Trust (Watson)*, 487 U.S. 977, 997 (1988).

³⁴² *Watson*, 487 U.S. at 991; *Wal-Mart*, 131 S.Ct. at 2254. The Court held in *Watson* that “disparate impact analysis is in principle no less applicable to subjective employment criteria than to objective or standardized tests. In either case, a facially neutral practice, adopted without discriminatory intent, may have effects that are indistinguishable from intentionally discriminatory practices.” *Watson*, 487 U.S. at 990. “If an employer’s undisciplined system of subjective decisionmaking has precisely the same effects as a system pervaded by impermissible intentional discrimination, it is difficult to see why Title VII’s proscription against discriminatory actions should not apply.” *Watson*, 487 U.S. at 990-991 (internal citation omitted).

³⁴³ 267 F.3d 147 (2d Cir. 2001) (internal citations omitted).

degree sufficient to reveal a causal relationship between the challenged practice and the disparity. Moreover, a plaintiff must show that each individual challenged employment practice has a significantly disparate impact.

If the plaintiffs succeed in their *prima facie* showing, the burden of persuasion then shifts to the employer to demonstrate one of two things. The first option is to challenge the plaintiffs' statistical proof. This may be done by introducing evidence to show that either no statistically significant disparity in fact exists or the challenged practice did not cause the disparity. To successfully contest the plaintiffs' statistical evidence, however, the employer has to convince the factfinder that its numerical picture is more accurate, valid, or reliable than the plaintiffs' evidence. If the employer is able to do so, it prevails and the case ends.

Assuming the employer is unable to successfully contest the plaintiffs' statistics, a second route is for the employer to demonstrate that the challenged practice or policy is job related for the position in question and consistent with business necessity. If the employer fails to demonstrate a business justification for the policy or practice, then the plaintiffs prevail. If the employer succeeds in establishing a business justification, however, the disparate impact claim proceeds to a third stage. At this third stage, the burden of persuasion shifts back to the plaintiffs to establish the availability of an alternative policy or practice that would also satisfy the asserted business necessity, but would do so without producing the disparate effect.

Should the plaintiffs succeed in establishing a Title VII disparate impact violation, the court may order declaratory and prospective class-wide injunctive relief. However, in order for an employee to obtain individual relief (*e.g.*, back pay), an individual inquiry is generally required in which each class member must show that he or she was among those adversely affected by the challenged policy or practice. If this showing is made, the class member is